

PAPER • OPEN ACCESS

## MEMS piezoelectric resonant microphone array for lung sound classification

To cite this article: Hai Liu *et al* 2023 *J. Micromech. Microeng.* **33** 044003

View the [article online](#) for updates and enhancements.

You may also like

- [A data-scalable randomized misfit approach for solving large-scale PDE-constrained inverse problems](#)  
E B Le, A Myers, T Bui-Thanh et al.
- [Computerized wheeze detection in young infants: comparison of signals from tracheal and chest wall sensors](#)  
Lia C Puder, Silke Wiltzki, Christoph Bühner et al.
- [Rubber muscle actuation with pressurized CO<sub>2</sub> from enzyme-catalyzed urea hydrolysis](#)  
Thomas M Sutter, Matthew B Dickerson, Terry S Creasy et al.

# MEMS piezoelectric resonant microphone array for lung sound classification

Hai Liu<sup>\*</sup>, Matin Berekatain<sup>1</sup>, Akash Roy<sup>1</sup>, Song Liu, Yunqi Cao, Yongkui Tang, Anton Shkel and Eun Sok Kim

Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, United States of America

E-mail: [hailiu@usc.edu](mailto:hailiu@usc.edu)

Received 20 October 2022, revised 12 February 2023

Accepted for publication 28 February 2023

Published 9 March 2023



## Abstract

This paper reports a highly sensitive piezoelectric microelectromechanical systems (MEMS) resonant microphone array (RMA) for detection and classification of wheezing in lung sounds. The RMA is composed of eight width-stepped cantilever resonant microphones with Mel-distributed resonance frequencies from 230 to 630 Hz, the main frequency range of wheezing. At the resonance frequencies, the unamplified sensitivities of the microphones in the RMA are between 86 and 265 mV Pa<sup>-1</sup>, while the signal-to-noise ratios (SNRs) for 1 Pa sound pressure are between 86.6 and 98.0 dBA. Over 200–650 Hz, the unamplified sensitivities are between 35 and 265 mV Pa<sup>-1</sup>, while the SNRs are between 79 and 98 dBA. Wheezing feature in lung sounds recorded by the RMA is more distinguishable than that recorded by a reference microphone with traditional flat sensitivity, and thus, the automatic classification accuracy of wheezing is higher with the lung sounds recorded by the RMA than with those by the reference microphone, when tested with deep learning algorithms on computer or with simple machine learning algorithms on low-power wireless chip set for wearable applications.

Keywords: piezoelectric MEMS microphone, resonant microphone array, acoustic transducer, lung sounds classification, wearable health sensor

(Some figures may appear in colour only in the online journal)

## 1. Introduction

About 7.4% of the world population suffer from chronic respiratory diseases, among which asthma and chronic obstructive pulmonary disease (COPD) are most common [1, 2]. As many as 262 million people are affected by asthma and more

than 1200 individuals die from asthma every single day on average [3]. Wheezing due to narrowing airway of the lung caused by asthma is a common symptom and can be sensed by a stethoscope [4–7]. Thus, lung sound monitoring can be very helpful for asthma patients, especially children, who cannot carry out the well-established pulmonary function tests accurately due to their inability to understand or follow the instruction on how to force air out of their lungs. Lung sounds can be monitored with electronic stethoscopes, but not for more than 1 h continuously due to their bulkiness and heaviness (stemming from the acoustic coupler needed to amplify faint lung sound) [8–10]. Also, a weak wheezing may be missed because of a low signal-to-noise ratio (SNR) of the microphone (used in the stethoscope). The published sensitivities

<sup>1</sup> These authors contributed equally.

\* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

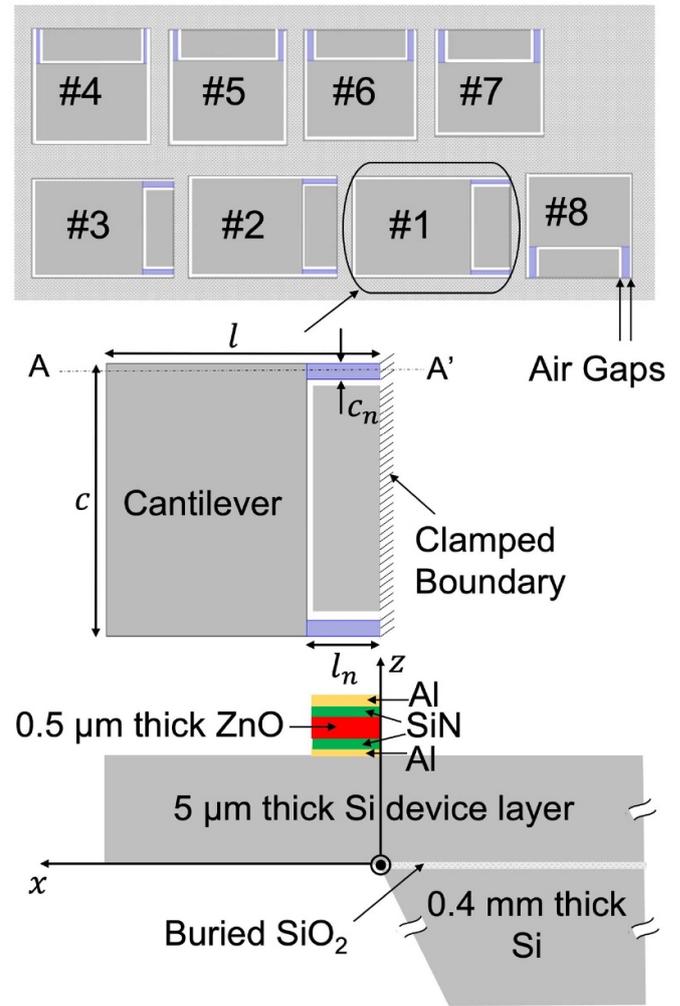
of commercial MEMS condenser microphones are between 5 (TDK INMP411) [11] and 25.12 mV Pa<sup>-1</sup> (TDK ICS-40730) [12], which depends on the applied bias voltage (typically about 10 V<sub>DC</sub>). Their SNRs for 1 Pa sound pressure are between 59 (Knowles SPH2430HR5H-B) [13] and 74 dBA (TDK ICS-40730) [12]. A MEMS condenser microphone with 22.39 mV Pa<sup>-1</sup> sensitivity and 73 dBA SNR over 22 Hz–22 kHz has been reported, but with 200 V<sub>DC</sub> applied bias voltage [14]. In the case of piezoelectric MEMS microphones, a bias voltage is not needed, and an unamplified sensitivity of 38 mV Pa<sup>-1</sup> over 100–700 Hz with the fundamental resonance at 890 Hz has been reported [15].

Microphone sensitivity is enhanced at the mechanical resonance of a microphone diaphragm when the resonance's quality factor ( $Q$ ) is greater than 1, and MEMS resonant microphones have been reported [16–23]. An array of such resonant microphones can mimic the human auditory system based on resonances of 30 000 cochlear hairs at the basilar membrane [24, 25]. A higher  $Q$  means a higher sensitivity at the resonance frequency, but over a narrower bandwidth. Thus, a diaphragm with multiple resonances or a resonant microphone array (RMA) consisting of multiple resonant microphones covering different frequencies is needed to cover a wide frequency range. A piezoelectric RMA with unamplified sensitivities of the resonant microphones being 34.6–131.4 mV Pa<sup>-1</sup> at their resonance frequencies between 169 and 662 Hz was reported for lung sound monitoring [21]. However, the noise floor and SNR of the RMA was not reported in [21].

This paper presents the design, fabrication, characterization, and application of a highly sensitive piezoelectric MEMS RMA for detection and automatic classification of wheezing in lung sounds. Measured unamplified sensitivity and SNR of the RMA are presented along with machine learning algorithms developed and implemented on a computer and on a commercial wireless chip set CYBLE-416045-02. Also presented are measured classification accuracies and speeds (directly related to energy consumption) for wheezing in lung sounds.

## 2. Design

Eight of the width-stepped Si cantilevers, with two narrow beams supporting a rectangular plate, are used for the resonant microphones in the RMA (figure 1). The resonance frequencies are Mel-spaced (denser at lower frequencies as humans are capable of distinguishing lower frequencies better) between 200 and 800 Hz (frequency range of wheezing). More cantilevers can cover more frequencies with highly sensitive resonances, albeit at the cost of larger size for the RMA. For the wheezing detection over 200–800 Hz, eight cantilevers offer good trade-off between the performance and the size. Piezoelectric thin film ZnO, which converts the cantilever bending stress (due to applied sound pressure) to voltage, is placed only over the narrow support beams for maximum average stress over a largest possible area. Electrical insulation layer SiN encapsulates ZnO to prevent charge transfer between



**Figure 1.** Resonant microphone array (RMA) schematic for lung sound classification with eight width-stepped cantilevers (top); top-view (middle) and cross-sectional view (bottom) of each width-stepped cantilever.

the top and bottom electrodes through ZnO for good sensitivity at low frequencies, as the resistivity of ZnO ( $10^7 \Omega \cdot \text{cm}$ ) is relatively low. The air gap between the cantilever and the Si base is as narrow as 20  $\mu\text{m}$  to minimize acoustic pressure leakage at low frequencies. The sizes of the cantilever resonant microphones are 3.6–2.3 mm (table 1), while the thickness of the Si cantilever is 5  $\mu\text{m}$  (table 2).

The fundamental resonance frequency of a width-stepped cantilever can be calculated through a beam free vibration equation for the cantilever displacements  $W_1(x)$  and  $W_2(x)$  (equation (1)) for the two parts having different widths

$$\begin{cases} \frac{d^4 W_1(x)}{dx^4} - \beta_1^4 W_1(x) = 0, & 0 \leq x \leq l_n \\ \frac{d^4 W_2(x)}{dx^4} - \beta_2^4 W_2(x) = 0, & l_n < x \leq l \end{cases} \quad (1)$$

where  $\beta_1^4 = \frac{m_1 \omega^2}{E_1 I_1}$  and  $\beta_2^4 = \frac{m_2 \omega^2}{E_2 I_2}$ , with  $m_1$  and  $m_2$  being the mass per unit length,  $E_1$  and  $E_2$  being the Young's modulus,

**Table 1.** Resonance frequencies and sizes of the resonant microphones in the RMA in figure 1.

No.	Resonance frequency $f$	Outline (mm)		Narrow part (mm)	
		Length $l$	Width $c$	Length $l_n$	Width $c_n$
1	253 Hz	3.6	2.3	0.9	0.13
2	308 Hz	3.4	2.3	0.8	0.14
3	367 Hz	3.2	2.2	0.7	0.15
4	429 Hz	2.6	2.6	0.8	0.115
5	495 Hz	2.6	2.6	0.7	0.15
6	565 Hz	2.5	2.5	0.7	0.17
7	639 Hz	2.4	2.4	0.75	0.18
8	717 Hz	2.3	2.3	0.7	0.185

**Table 2.** The layer thicknesses of a width-stepped cantilever resonant microphone in figure 1.

Layer #	Name	Thickness $h$
1	Si device layer	5 $\mu\text{m}$
2	Al ground electrode	0.2 $\mu\text{m}$
3	SiN insulator	0.1 $\mu\text{m}$
4	Piezoelectric ZnO	0.5 $\mu\text{m}$
5	SiN insulator	0.1 $\mu\text{m}$
6	Al top electrode	0.2 $\mu\text{m}$

and  $I_1$  and  $I_2$  being the moment of inertia, at part 1 and 2, respectively. Once  $\beta_1$  and  $\beta_2$  are solved through equation (1) [26], the resonance frequency  $f$  can be obtained as follows

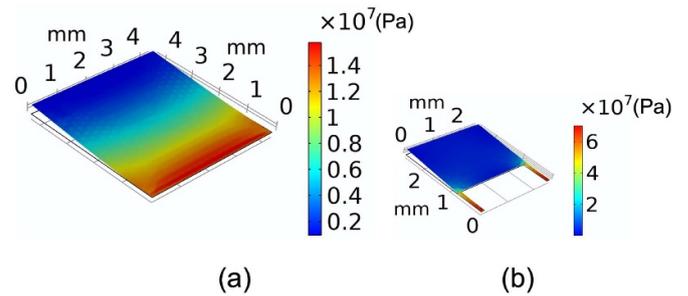
$$f = \frac{\beta_1^2}{2\pi} \sqrt{\frac{E_1 I_1}{m_1}} = \frac{\beta_2^2}{2\pi} \sqrt{\frac{E_2 I_2}{m_2}}. \quad (2)$$

Compared to a rectangular cantilever with one fixed and three free ends (figure 2(a)), the bending-induced stress due to an applied pressure and the cantilever size for a same fundamental frequency are higher and smaller, respectively, for a width-stepped cantilever (figures 2–4). Thus, a higher unamplified sensitivity is expected with a piezoelectric microphone built on a width-stepped cantilever than that on a standard cantilever, as a piezoelectric film ZnO is placed only on the support beams of a width-stepped cantilever (figure 1). The voltage  $V$  produced across the ZnO thickness due to average stress  $\sigma$  induced by bending caused by an applied pressure is

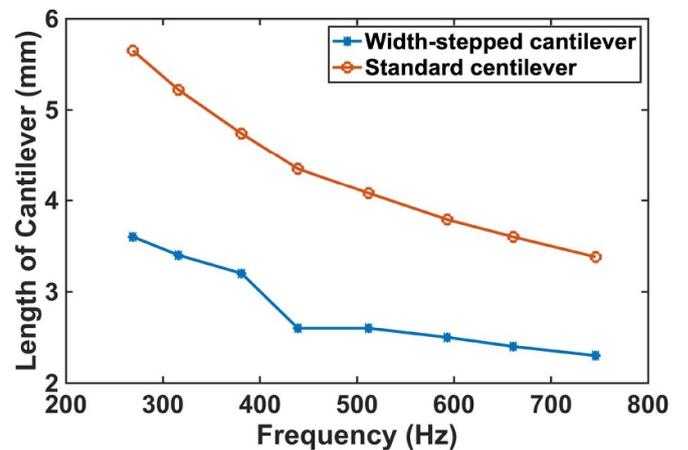
$$V = \frac{\sigma \times d_{31} \times A}{C} = \frac{\sigma \times d_{31} \times A}{\frac{\epsilon_0 \epsilon_r A}{t}} = \frac{\sigma \times d_{31} \times t}{\epsilon_0 \epsilon_r} \quad (3)$$

where  $C$ ,  $A$  and  $t$  are the ZnO’s capacitance, area, and thickness, respectively, while  $d_{31}$  and  $\epsilon_r$  are the piezoelectric coefficient and relative permittivity of the piezoelectric film, respectively, with  $\epsilon_0$  being vacuum permittivity.

The resonance frequency of a width-stepped cantilever is not only dependent on the size of the whole cantilever but also on the size of the narrow segment as shown in table 1. Therefore, there is more design flexibility for the width-stepped cantilever. To make the RMA illustrated in figure 1 smaller, the

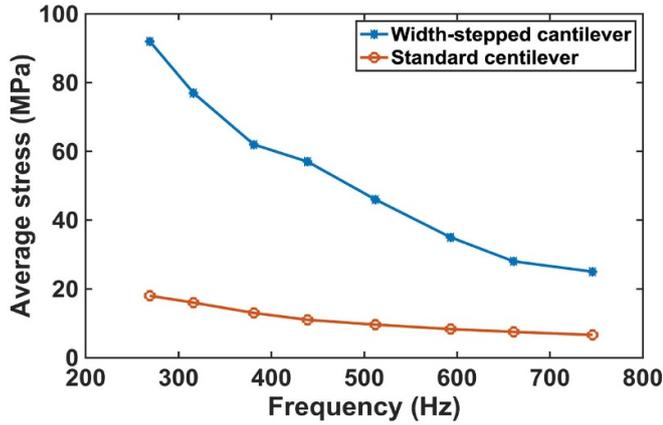


**Figure 2.** Finite-element-analysis (FEA) simulated stresses of 5  $\mu\text{m}$  thick cantilevers having the same fundamental resonant frequency of 436 Hz under 1 Pa pressure; (a) a standard cantilever of  $4.35 \times 4.35 \text{ mm}^2$  showing 11 MPa average stress near the fixed end and (b) a width-stepped cantilever of  $2.6 \times 2.6 \text{ mm}^2$  with 57 MPa average stress near the fixed ends of the two supporting beams.

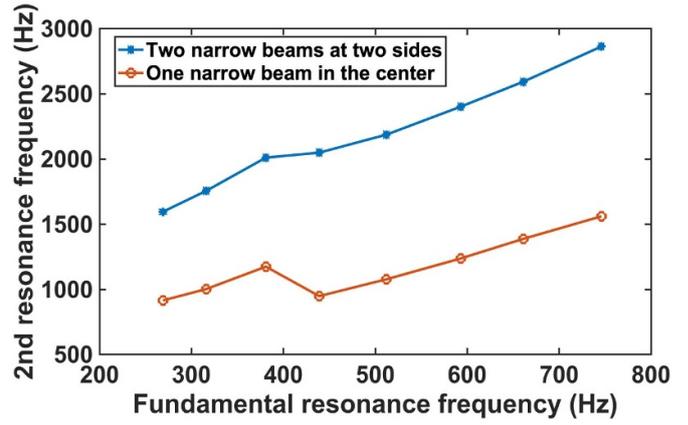


**Figure 3.** Calculated lengths of a width-stepped cantilever (table 1) and a standard cantilever vs fundamental resonance frequency.

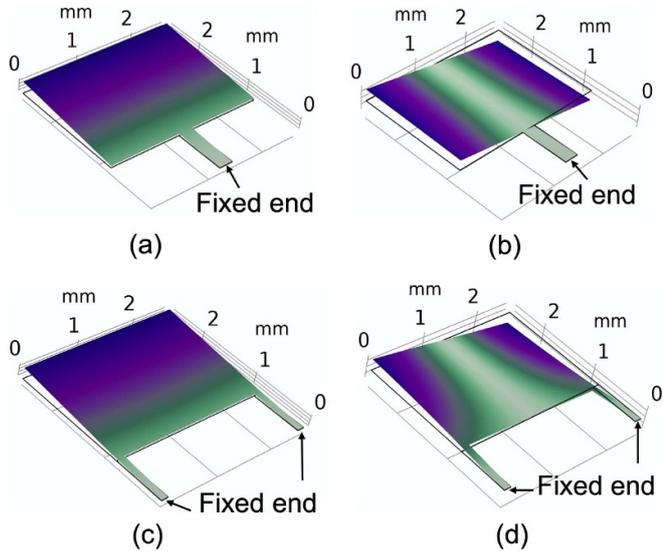
width-stepped cantilever #4 (having the fundamental resonance frequency of 429 Hz) in the array is designed to have the same length  $l$  as #5 (having the fundamental resonance frequency of 495 Hz) but with longer and narrower Narrow Part (table 1). This is why the curves of the length, average stress, and 2nd resonance frequency are not smooth for the width-stepped cantilever at #4 in figures 3, 4 and 6, respectively.



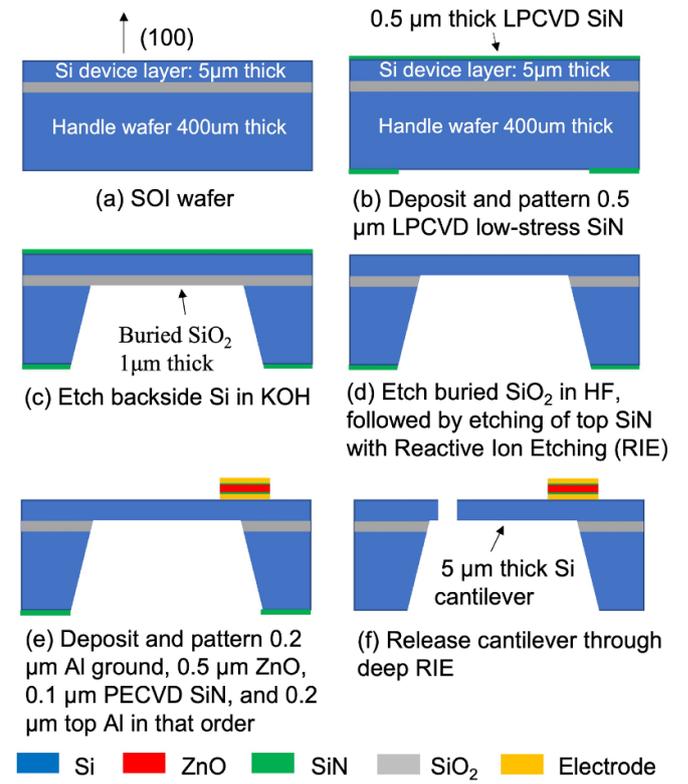
**Figure 4.** Calculated average stresses near the fixed ends for the width-stepped cantilever (table 1) and standard cantilever under 1 Pa sound pressure vs fundamental resonance frequency.



**Figure 6.** Simulated the 2nd resonance frequencies of the two types of width-stepped cantilevers vs fundamental resonance frequency. For each fundamental resonance frequency, the total length and width of the two types of the cantilevers are the same. The width of the one narrow beam in the center is twice of the beam width at the ends, while the length is the same.



**Figure 5.** Simulated vibrations of width-stepped cantilevers at (a) and (b) the fundamental (436 Hz) and second-harmonic (949 Hz) resonance frequencies, respectively, with one narrow support beam in the center, (c) and (d) the fundamental (436 Hz) and second-harmonic (2049 Hz) resonance frequencies, respectively, with two narrow support beams at the two ends. The width of the one narrow beam in the center is twice that of the narrow beam at the end, while the length is the same. The total width and length of the cantilevers are the same to be  $2.6 \times 2.6 \text{ mm}^2$ .

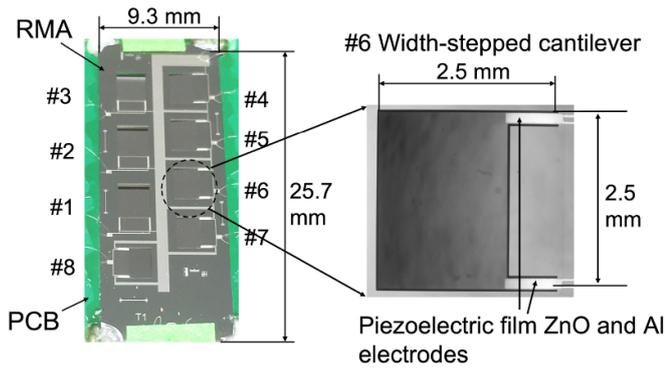


**Figure 7.** Brief fabrication process of the RMA.

A width-stepped cantilever with one narrow support beam in the center (figure 5(a)) has the 2nd harmonic resonance frequency close to the fundamental resonance frequency (figures 5(b) and 6), which may result in the 2nd resonance overlapping the fundamental resonance of another cantilever in an RMA. As we would like to utilize the fundamental resonance of each resonant microphone in an RMA and avoid any interference of the harmonics, a width-stepped cantilever with two narrow support beams (figures 5(c), (d) and 6) is designed for each resonant microphone in the RMA.

### 3. Fabrication

The RMA is fabricated on a silicon-on-insulator wafer with 5 μm thick Si device layer (figure 7(a)). First, 0.5 μm thick low-stress SiN is deposited with low pressure chemical vapor deposition and patterned (figure 7(b)) for etch mask during KOH etching the Si (figure 7(c)). After etching the buried SiO<sub>2</sub> in buffered HF, followed by etching of the top SiN in reactive ion etching (figure 7(d)), we sputter-deposit and pattern



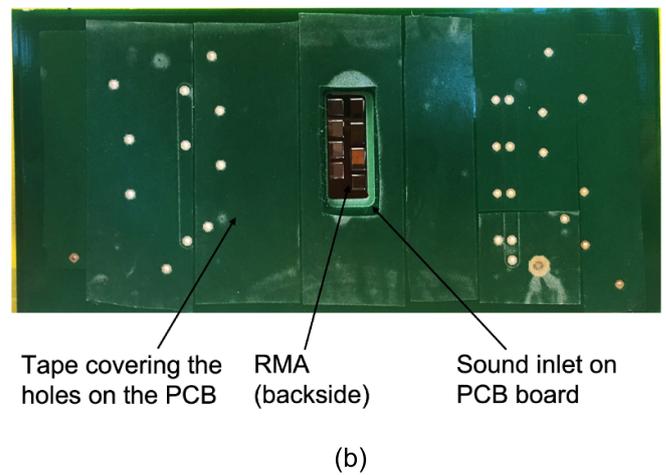
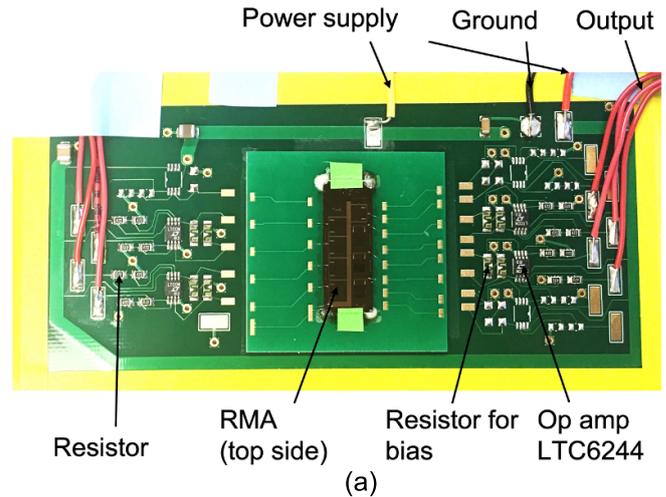
**Figure 8.** Photos of a fabricated RMA with eight width-stepped cantilever resonant microphones.

**Table 3.** Measured residual stresses of thin films in the cantilever.

Film	Stress (MPa)
Sputter-deposited ZnO	-350
Sputter-deposited Al	-30
PECVD SiN	-50

**Table 4.** Measured paralleled capacitance and resistance of the resonant microphones in the RMA.

Microphone #	Capacitance	Resistance
1	23.9 pF	>10 GΩ
2	25.4 pF	>10 GΩ
3	23.5 pF	>10 GΩ
4	24 pF	>10 GΩ
5	26.3 pF	>10 GΩ
6	27.6 pF	>10 GΩ
7	28.5 pF	>10 GΩ
8	27.2 pF	>10 GΩ



**Figure 9.** Top-side (a) and bottom-side (b) photos of the RMA on a printed circuit board (PCB) with op-amp based circuits.

Al and piezoelectric ZnO, deposit SiN with plasma-enhanced chemical vapor deposition (PECVD) and pattern SiN, and then sputter-deposit and pattern Al (figure 7(e)). After dicing RMAs from the wafer, the cantilevers are released on each chip through etching Si on the diaphragms of each RMA (figure 7(f)).

Long cantilevers (particularly, #1, #2 and #3) in the fabricated RMA (figure 8) show substantial downward warpage due to relatively large compressive residual stress in the ZnO film (table 3). The residual stresses of the thin films in table 3 are the average values calculated through measuring the curvature of a 3" wafer by a profilometer DektakXT before and after the film deposition.

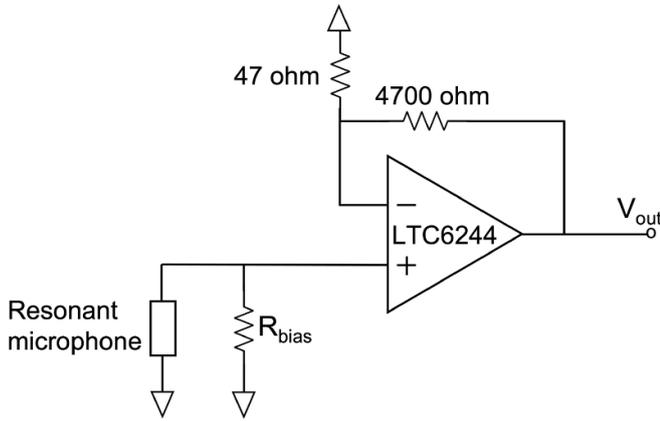
## 4. Characterization

### 4.1. Unamplified sensitivity

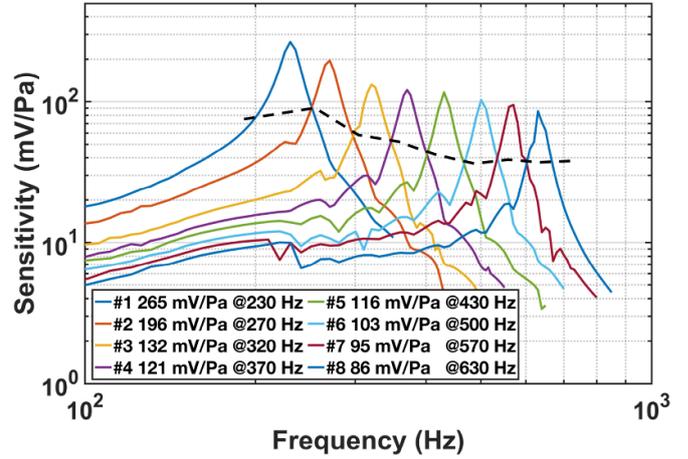
The measured capacitances and resistances of the resonant microphones (table 4) are close to the designed values.

The fabricated RMA is placed over a slot (for sound input) in a printed circuit board (PCB) (figure 9) and voltage amplifiers based on LTC6244 op amp with input resistance and capacitance of  $10^{12} \Omega$  and 2.1 pF, respectively (figure 10). The signal from each resonant microphone of the RMA is magnified and recorded separately without connecting to each other. The measured sensitivity from the amplifier output is divided by the amplification factor of 101 for unamplified sensitivity. A bias resistor of 1 GΩ (figure 10) is used for DC-biasing the op amp without affecting the low frequency response of the piezoelectric microphone.

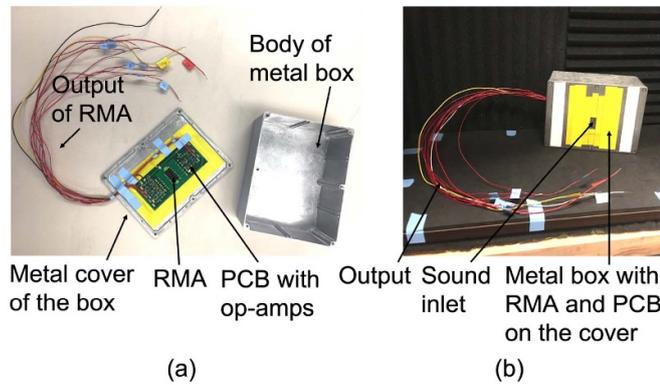
The PCB is mounted to a cover plate (with a slot for sound input) of a metal box which blocks electromagnetic interference (figure 11). The outputs of the microphone amplifiers are connected to a data acquisition system (ROGA Plug.n.DAQ). The sound input to the resonant microphone is calibrated with a reference measurement microphone (GRAS 40AO, noise floor 25 dBA, sensitivity  $12.5 \text{ mV Pa}^{-1}$ , bandwidth 3.15 Hz–20 kHz) with both the RMA and the reference microphone



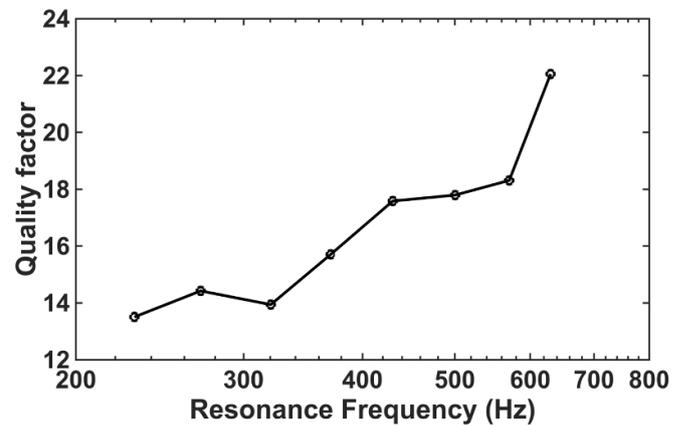
**Figure 10.** Schematic of non-inverting op-amp-based amplification circuit for each piezoelectric resonant microphone in the RMA.



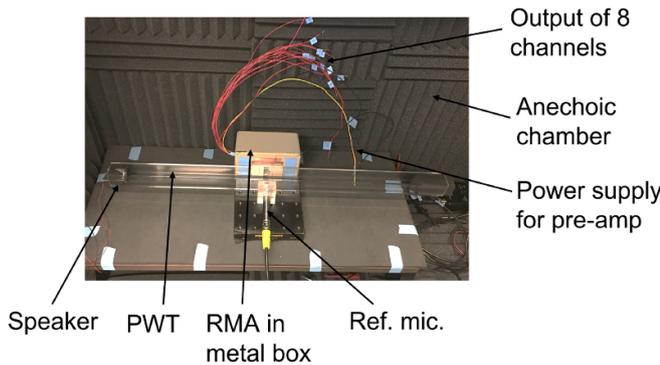
**Figure 13.** Measured unamplified sensitivities of the eight resonant microphones in the RMA. It is above  $35 \text{ mV Pa}^{-1}$  for the whole RMA between 200 Hz and 650 Hz as indicated by the black dash line.



**Figure 11.** Photos of (a) the RMA on a PCB in a metal box and (b) the packaged RMA in an anechoic chamber.



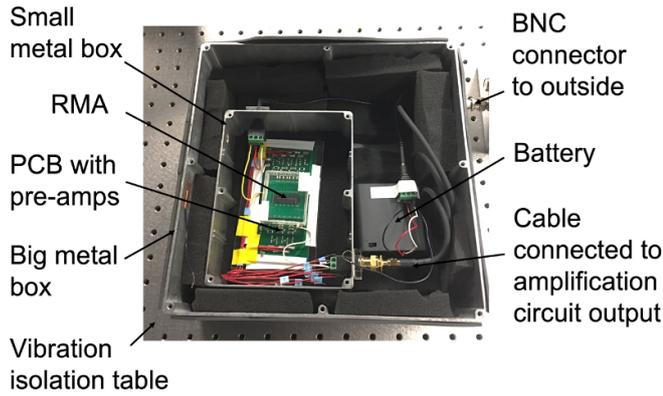
**Figure 14.** Measured quality factors of the resonant microphones in the RMA.



**Figure 12.** Photo of the microphone sensitivity measurement setup with a reference microphone GRAS 40AO across a plane wave tube (PWT), the cross section of which is  $25.4 \times 25.4 \text{ mm}^2$ , in an anechoic chamber.

being placed next to each other in a plane wave tube (PWT) in an anechoic chamber (figure 12). A loudspeaker placed at one end of the PWT delivers same sound pressure to the RMA and the reference microphone which are located near to each other.

The measured unamplified sensitivities of the eight resonant microphones in the RMA are as high as  $265\text{--}86 \text{ mV Pa}^{-1}$  at the eight resonance frequencies (figure 13). The sensitivity of the RMA at all frequencies between 200 and 650 Hz is above  $35 \text{ mV Pa}^{-1}$  (above the dash line in figure 13). The measured resonance frequencies are lower than the designed ones (table 1), mainly because the Si cantilevers turn out to be  $4 \mu\text{m}$  thick, rather than  $5 \mu\text{m}$  thick. The sensitivity curve of each resonant microphone shows ripples near the resonance frequencies of the other resonant microphones due to electrical crosstalk among the resonant microphones in the RMA. This phenomenon can be reduced through better grounding of the microphones and amplification circuits. The quality factors (based on resonant frequency  $f_0$  divided by the  $-3 \text{ dB}$  bandwidth in figure 13) of the resonant microphones are between 13.5 and 22 (figure 14). The damping coefficient of a smaller resonant microphone is usually smaller because the damping is mainly from the air surrounding the cantilever.



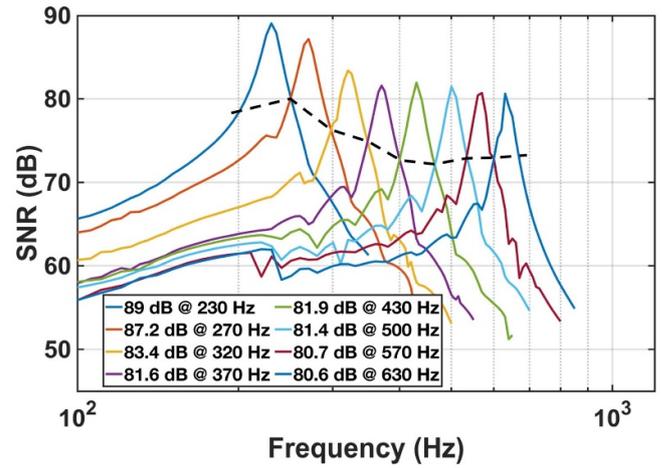
**Figure 15.** Photo of noise measurement setup with the RMA in a double metal box on an optical table for isolation from electromagnetic noise, acoustic noise, and mechanical vibration. The top of the metal box is covered during noise measurement.

Consequently, the quality factor is usually higher for the resonant microphones with higher resonance frequencies and smaller size. Higher quality factor leads to higher unamplified sensitivity and lower noise floor of a resonant microphone while the bandwidth (over which the sensitivity is enhanced by the resonance) is narrower. Therefore, we need proper quality factors for the resonant microphones in the RMA so that the RMA has both high sensitivities to detect weak sound and enough bandwidth to cover the frequency range of interest.

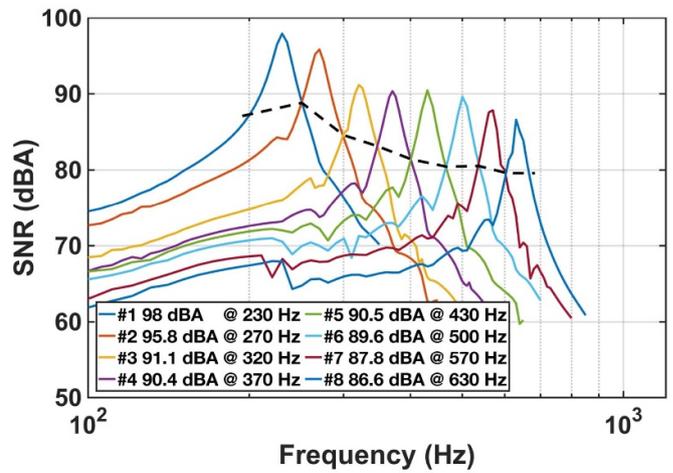
#### 4.2. Noise floor and SNR

To measure the noise without any electromagnetic or sound interference noise, the RMA and amplification circuit are placed in a double metal box with battery (figure 15).

With the double metal box on a vibration isolation table, the output of the amplification circuit for each resonant microphone in the RMA is divided by the amplification (101) for input-referred noise of each resonant microphone and its amplifier. The measured input-referred root-mean-square (RMS) noise over 20 Hz–20 kHz observation bandwidth is 8–10 and 3–4  $\mu\text{V}$  before and after A-weighting, respectively. The noise floor in pressure is obtained by dividing the input-referred RMS noise voltage by the unamplified sensitivity, while the noise floor in dB is  $20 \log(\text{noise-floor-in-Pa}/\text{reference-pressure})$  where the reference pressure is  $2 \times 10^{-5}$  Pa. And the SNR for 1 Pa sound pressure input is obtained by deducting the noise floor in dB from 94 dB, as 1 Pa sound pressure is 94 dB ( $= 20 \log(1/(2 \times 10^{-5}))$ ). The measured SNRs of the resonant microphones at their resonances for 1 Pa sound pressure are 89.0–80.6 dB before A-weighting and 98.0–86.6 dBA after A-weighting (figure 16). Over all the frequencies between 200 and 650 Hz where wheezing in lung sounds is prominent, the SNR are 89.0–73.0 dB before A-weighting and 98.0–79.0 dBA after A-weighting (above the black dash line in figure 16). The spectral noise density of all resonant microphones in the RMA without A-weighting are lower than  $1 \mu\text{V}/\sqrt{\text{Hz}}$  (figure 17) with the  $1/f$  noise of the op amp dominating, indicating that



(a)



(b)

**Figure 16.** Measured signal-to-noise ratios (SNRs) of the resonant microphones in the RMA at 1 Pa sound pressure vs frequency (a) without A-weighting and (b) with A-weighting. The SNR of the whole array is above the black dash line.

the external sound and vibration are isolated very well during the noise measurement.

## 5. Lung sounds detection and classification

Well-annotated lung sounds from International Conference on Biomedical and Health Informatics (ICBHI) Respiratory Sound Database [27] are played by a loudspeaker and recorded with the RMA and the reference microphone in a set-up shown in figure 12. The recordings are analyzed in time and frequency domains, and processed through deep learning and machine learning algorithms for automatic classification of wheezing in the lung sounds, to show the advantages of the RMA over a standard microphone.

### 5.1. Recorded signals in time and frequency domain

Wheezing in lung sounds is easily recognizable in the recording by the RMA in time (figure 18(a)) and spectrogram

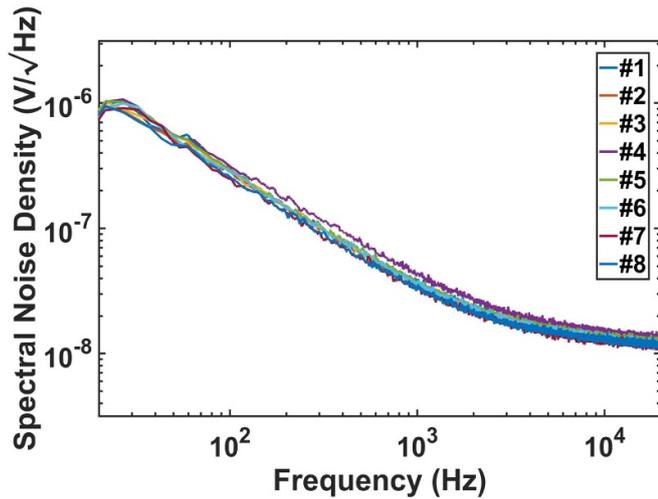


Figure 17. Measured noise spectral density of the resonant microphones in the RMA without A-weighting.

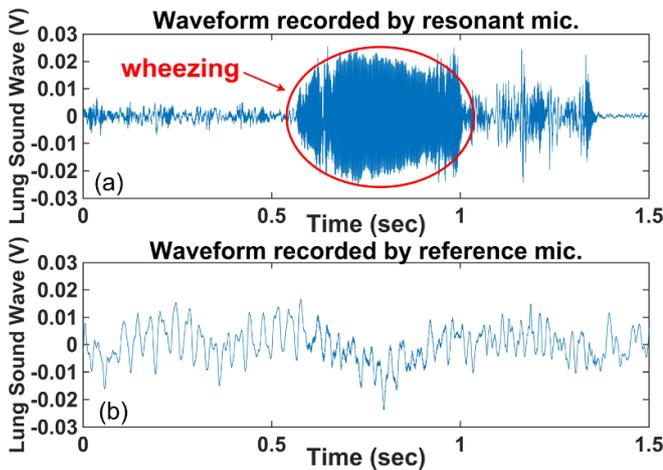


Figure 18. Waveform of a lung sound with strong wheezing recorded by (a) the resonant microphone #7 in the RMA (which records distinguishable wheezing) and (b) the reference microphone GRAS 40AO (which does not record distinguishable wheezing).

(figure 19(a)), while the recording by the reference microphone shows little wheezing feature in time (figure 18(b)) and a weak feature in the spectrogram (figure 19(b)). A weak wheezing is not visible in the time recordings by the RMA and the reference microphone (figures 20(a) and 21(a)). Such a weak wheezing, though, can still be distinguished in the spectrogram of the recording by the RMA (figure 20(b)), but is not distinguishable in the spectrogram of the recording by the reference microphone (figure 21(b)). Thus, wheezing in the lung sounds can be recognized better in both time and frequency domain by the RMA than that by the conventional microphone.

### 5.2. Automatic wheeze classification with deep learning

Fifty lung sounds from the ICBHI database [27], with twenty-five of them having wheezing, are played by a loudspeaker

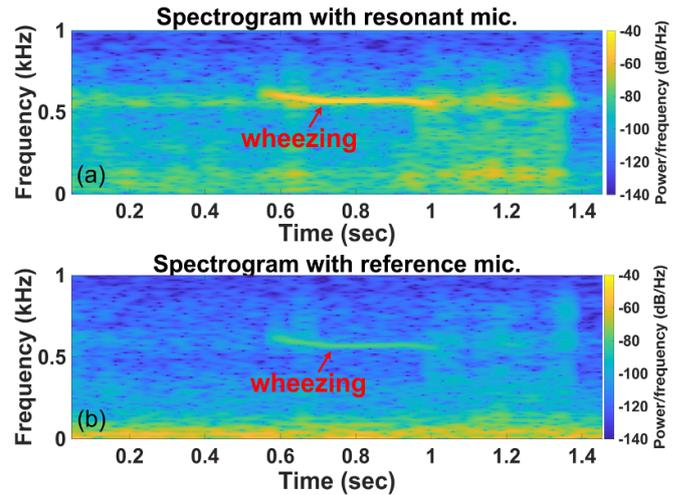


Figure 19. Spectrogram of a lung sound with strong wheezing recorded by (a) the resonant microphone #7 in the RMA (wheezing is distinguishable) and (b) the reference microphone (wheezing is distinguishable but not as obvious as the one recorded by the resonant microphone).

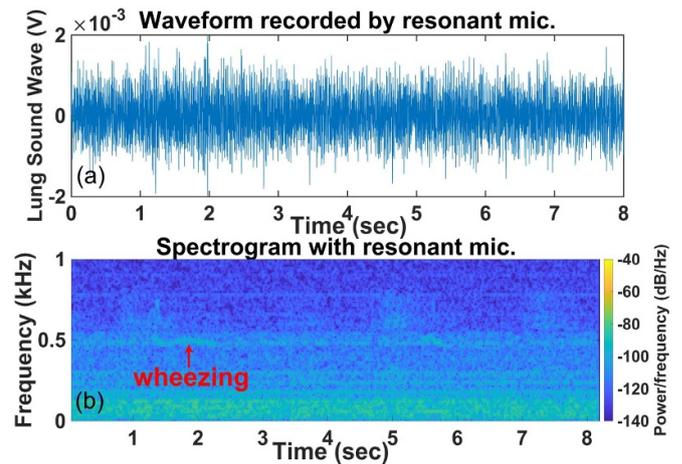
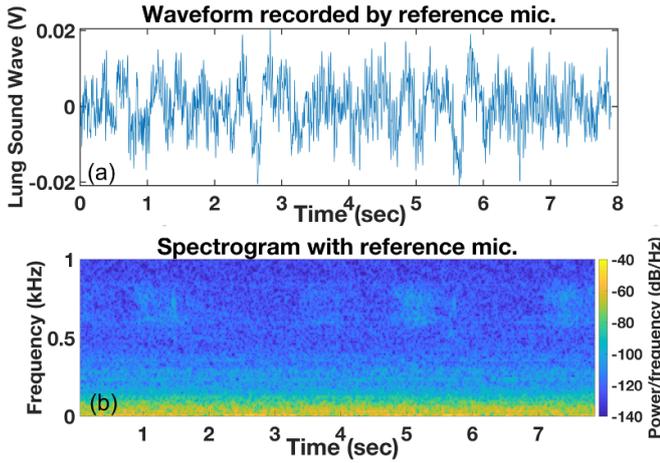


Figure 20. Waveform (a) and spectrogram (b) of a lung sound with weak wheezing recorded by the resonant microphone #6 in the RMA vs time; wheezing is not distinguishable in the waveform, while wheezing is distinguishable in the spectrogram.

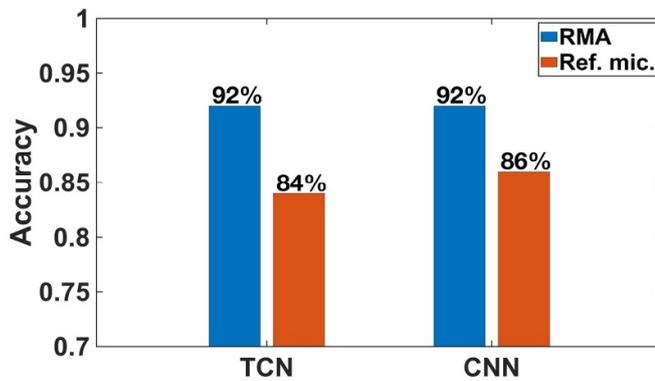
and recorded by the RMA and the reference microphone (figure 12). The recordings are classified by deep learning algorithms, and the classification accuracies are compared.

With temporal convolutional networks (TCNs) [28], the recorded lung sounds in time domain are processed without any pre-processing for the classification. Twelve-layer networks are used in TCN to extract a ten-dimensional feature vector for the classification. On the other hand, for convolutional neural networks (CNNs), pre-extracted Mel-frequency cepstral coefficients are used.

K-fold cross-validation is applied so that all the data can be used for both training and test. The recordings are divided into  $K$  (5 in this case) groups randomly. One group is used for the test while the other groups are used for the training at



**Figure 21.** Waveform (a) and spectrogram (b) of a lung sound with weak wheezing recorded by the reference microphone; wheezing is not distinguishable.



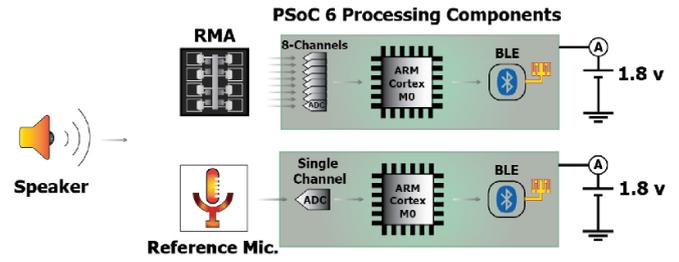
**Figure 22.** Wheezing automatic classification accuracy for lung sounds recorded by the RMA and the reference microphone and processed by two deep learning algorithms TCN and CNN.

each iteration until every group has been tested. The classification accuracy is the average accuracy of all the  $K$  iterations. With  $K$  being 5, 40 recordings are used for the training, and 10 recordings are used for the test in each iteration, for a total of 5 iterations. The classification accuracies of the lung sounds recorded by the RMA with both TCN and CNN are higher than what are obtained with the reference microphone (figure 22).

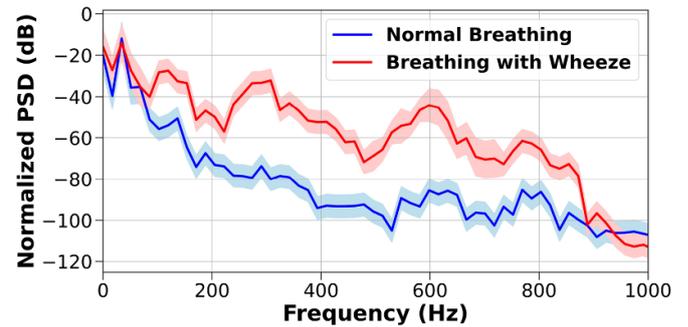
### 5.3. Automatic classification with machine learning on a chip set for wearable wireless communication

As deep learning algorithms cannot be implemented on a low power chip set for wireless communication such as Infineon CYBLE-416045-02 [29], which contains a microcontroller unit (MCU) and other components including antenna. We have developed and tested machine learning algorithms on the MCU (PSoC 63), which contains analog-to-digital converters, central processing unit, memory and blue tooth low energy transceiver (figure 23).

Each of the recordings is divided into 7650 pieces with each piece being 40 ms long for feature extraction. Mel-Spectrum features are extracted for classification as the frequency



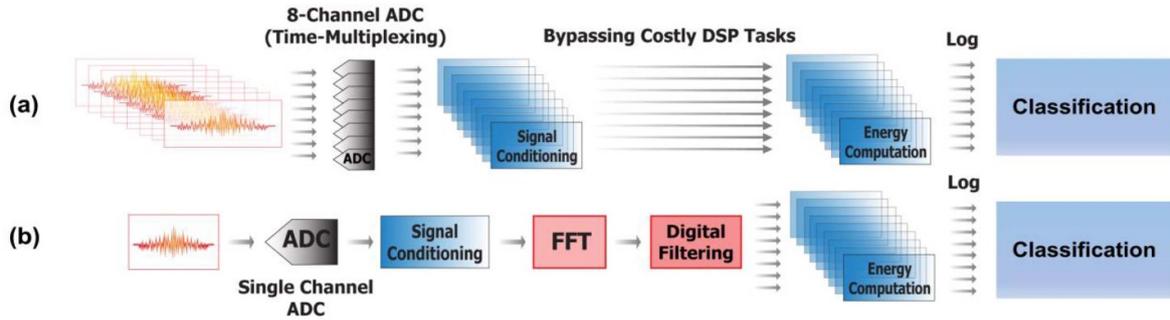
**Figure 23.** Schematic of the setups for audio recordings (by the RMA and the reference microphone) and processing with the wireless chip set (for feature extraction, classification through machine learning, and Bluetooth communication) to compare the performances of the RMA (top) and the reference microphone (bottom) for wheezing classification in lung sounds.



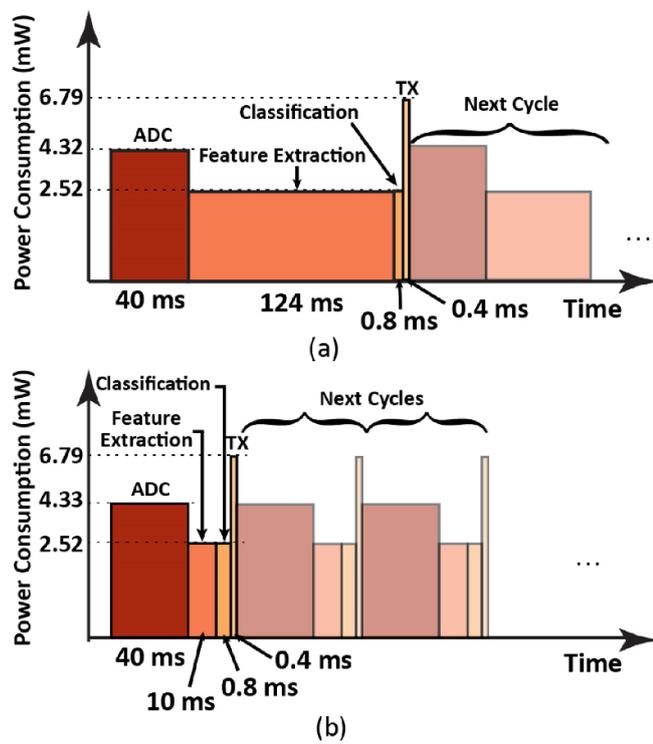
**Figure 24.** Spectral signature averaged per frame for lung sounds with and without wheezing recorded. The shaded regions indicate the standard deviation at each frequency. The power spectral density (PSD) of the lung sounds with wheezing is higher than that without wheezing especially between 300 and 600 Hz.

spectra of the lung sounds with and without wheezing are quite different (figure 24). On the data recorded by the reference microphone, fast Fourier transform (FFT) and digital filtering are applied for the feature extraction (figure 25(b)). However, with the data recorded by the RMA, the features at different frequencies are obtained through calculating the energy at each recording by individual resonant microphone in the RMA with its unique capability of acoustically filtering the audio signal (figure 25(a)). Thus, the feature extraction with the RMA is much faster (more than ten times) than that with the reference microphone as FFT is time consuming (figure 26). Though the idea of calculating the energy from individual channel of an RMA without FFT has been reported [30], the stronger MCU in PSoC 63 and optimized algorithms have resulted much faster signal processing.

Two machine models, Gaussian Naïve Bayes and support vector machine, are developed for the classification based on the extracted features. These classifiers are trained for single frame prediction at each moment because temporal-variation and multi-frame analyses require more memory and processing speed from PSoC 63. The recorded data are split with 70% for training and 30% for testing. The training is implemented on a desktop computer, and then the parameters of the algorithms obtained from the training are transferred to PSoC 63 for the test. We have measured the classification accuracy which is equal to  $(t_p + t_n) / (t_p + f_p + t_n + f_n)$  and F1



**Figure 25.** Block diagram illustrating the steps for the computing Mel-spectrogram features with (a) the eight-channel RMA and (b) the conventional approach using a traditional microphone with a digital filter bank. The FFT and multiple digital triangular filtering in (b) is not needed for (a).

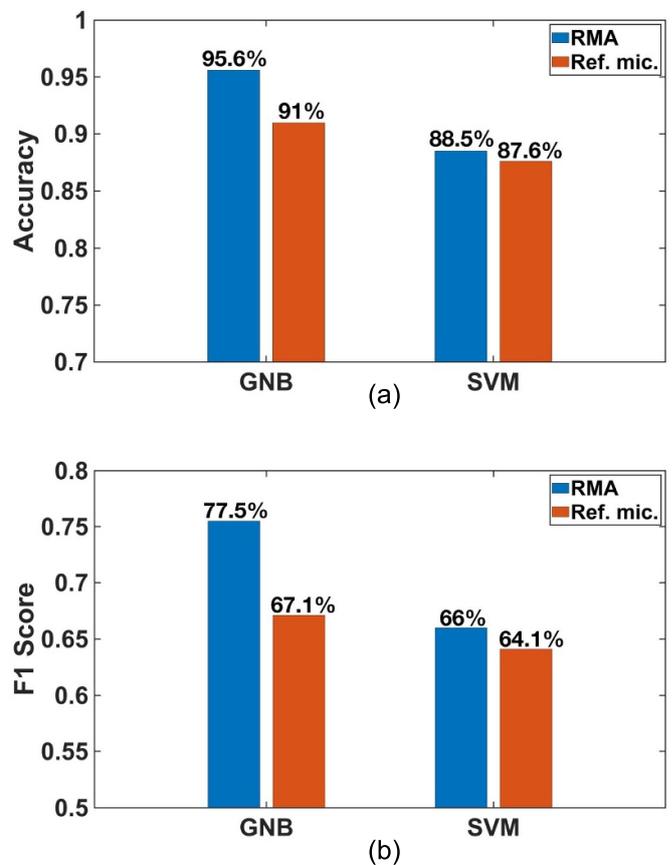


**Figure 26.** Power consumption vs time for sequential steps of automatic wheezing detection (a) with the reference microphone and (b) with the RMA. The feature extraction process by the RMA is  $\sim 12$  times faster and consumes 92% energy less than that by the reference microphone.

score which is equal to  $2PR/(P+R)$ , with  $t_p \equiv$  true positive,  $t_n \equiv$  true negative,  $f_p \equiv$  false positive,  $f_n \equiv$  false negative,  $P \equiv t_p/(t_p+f_p)$  and  $R \equiv t_p/(t_p+f_n)$ . Both the classification accuracy and F1 score are better with the recordings with RMA than with the reference microphone (figure 27).

### 6. Discussion

The unamplified sensitivities of the RMA ( $265\text{--}35\text{ mV Pa}^{-1}$ ) over 200–650 Hz where wheezing is prominent are higher than other MEMS microphones reported (figure 28), albeit with a larger size. The noise floor of the RMA also is lower than other



**Figure 27.** (a) Accuracies and (b) F1 scores of the classifications based on Gaussian Naïve Bayes (GNB) and support vector machine (SVM) algorithms for lung sounds recorded by the RMA and the reference microphone.

reported MEMS microphones over 200–650 Hz (figure 29). Thus, the minimum detectable wheezing signature in lung sounds is better with the RMA. If a microphone is targeted for a limited frequency range, an RMA with multiple resonances over the frequency range is shown to offer unprecedented minimum detectable sound level. Although the sensitivity and noise floor of the RMA is not flat, we did not find the effect of this un-flatness on the lung sound classification.

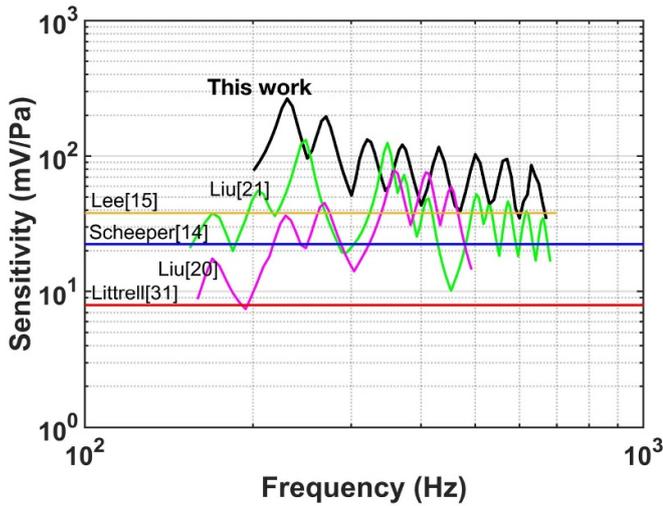


Figure 28. Comparison of the unamplified sensitivities of the reported MEMS microphones vs frequency.

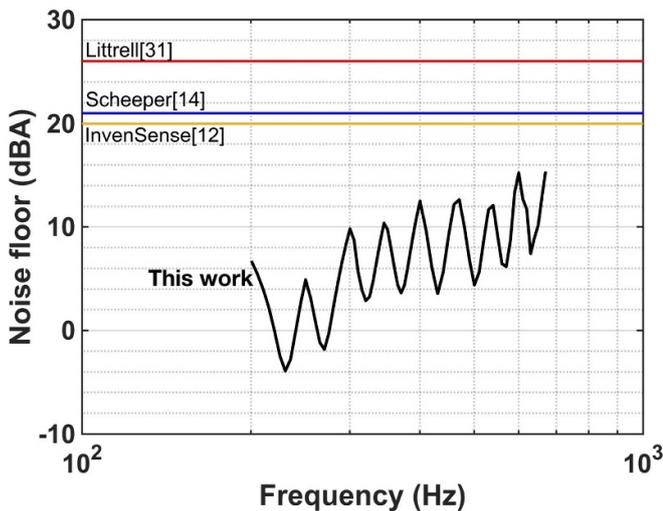


Figure 29. Comparison of the noise floors of the reported MEMS microphones vs frequency.

Design innovation and optimization are important to make the resonant microphones small and highly sensitive. The width-stepped cantilever design with two narrow beams supporting a rectangular plate is shown to be smaller and to offer higher sensitivity than a standard cantilever (figures 2–4) which has much less bending stiffness than a diaphragm with its four edges clamped. The size can be reduced further with a spiral structure or cantilever with serpentine support beams [20, 21], which has exhibited less sensitivity than the current design in this paper.

With the conventional approach, the number of digital filters can be increased to improve the classification accuracy (figure 30), but at increased process time and power consumption. The accuracy and F1 score with reference microphone plus 40 filters are still lower than that with the proposed RMA.

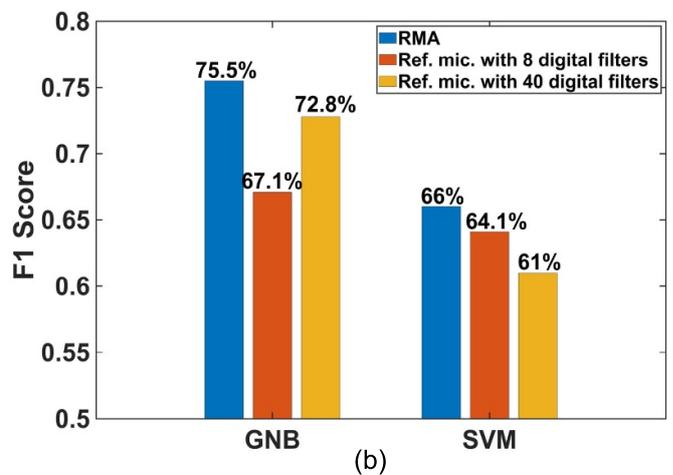
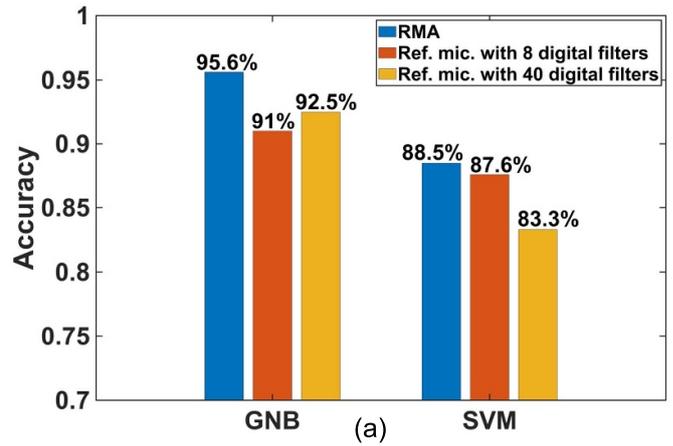


Figure 30. (a) Accuracies and (b) F1 scores for automatic classification of wheezing for lung sounds with the RMA without digital filters and the reference microphone with 8 and 40 digital filters.

Therefore, more advanced algorithms with poor quality data from microphones with higher noise floor may not compete with simple algorithms with high quality data from the RMA with extremely low noise floor. Furthermore, the signal processing is much faster (figure 26) with the RMA which inherently filters sounds into specific bandwidths. Thus, the current work shows the significant advantages of the RMA for real-time lung sound monitoring and classification with a wearable stethoscope.

## 7. Summary

An array of piezoelectric MEMS resonant microphones, with novel width-stepped cantilever design with two narrow beams supporting a rectangular plate, has been developed with Mel-distributed resonance frequencies to cover the frequency range where wheezing in lung sounds is prominent, and is shown to have the highest unamplified sensitivity and SNR in this

frequency range compared with other reported MEMS microphones. With the array, wheezing in lung sound is shown to be detected and automatically classified better than with a reference microphone. The automatic classification accuracies for wheezing are higher with the RMA for both deep learning (performed on a computer) and machine learning (performed on a chip set for wearable wireless communication). In addition, the signal processing with the RMA is shown to be more than ten times faster and consumes 92% less energy than that with a traditional microphone on a low power chip set for wearable wireless communication. Therefore, the current work paves the way for a wearable stethoscope to continuously monitor and automatically classify lung sounds in real-time so that patients or caregivers may be alerted and also so that medical professionals may have recordings of relevant lung sounds.

### Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

### Acknowledgments

This paper is based on the work supported by National Institute of Health under R21DC016468 and 1R01HL165138.

### ORCID iDs

Hai Liu  <https://orcid.org/0000-0003-4461-9748>

Yongkui Tang  <https://orcid.org/0000-0003-0817-4956>

### References

- [1] Soriano J B *et al* 2020 Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017 *Lancet Respir. Med.* **8** 585–96
- [2] Ferkol T and Schraufnagel D 2014 The global burden of respiratory disease *Ann. Am. Thorac. Soc.* **11** 404–6
- [3] World Health Organization 2022 Asthma (available at: [www.who.int/news-room/fact-sheets/detail/asthma](http://www.who.int/news-room/fact-sheets/detail/asthma)) (Accessed 5 October 2022)
- [4] Gong H 1990 Wheezing and asthma *Clinical Methods: The History, Physical, and Laboratory Examinations* 3rd edn, ed H K Walker, W D Hall and J W Hurst (Boston: Butterworths) ch 37
- [5] Castro-Rodríguez J A, Holberg C J, Wright A L and Martinez F D 2000 A clinical index to define risk of asthma in young children with recurrent wheezing *Am. J. Respir. Crit. Care Med.* **162** 1403–6
- [6] Meslier N, Charbonneau G and Racineux J L 1995 Wheezes *Eur. Respir. J.* **8** 1942–8
- [7] Priftis K N, Hadjileontiadis L J and Everard M L (eds) 2018 *Breath Sounds: From Basic Science to Clinical Practice* (Berlin: Springer)
- [8] Eko 2022 Electronic & digital stethoscopes for sale (available at: <https://shop.ekohealth.com/>) (Accessed 8 October 2022)
- [9] Thinklabs 2022 Thinklabs one digital stethoscope (available at: [www.thinklabs.com](http://www.thinklabs.com)) (Accessed 7 October 2022)
- [10] Li S H, Lin B S, Tsai C H, Yang C T and Lin B S 2017 Design of wearable breathing sound monitoring system for real-time wheeze detection *Sensors* **17** 171
- [11] TDK InvenSense 2014 Datasheet of INMP411 (available at: <https://invensense.tdk.com/download-pdf/inmp411-datasheet/>) (Accessed 8 October 2022)
- [12] TDK InvenSense 2020 Datasheet of ICS-40730 (available at: <https://invensense.tdk.com/download-pdf/ics-40730-datasheet/>) (Accessed 8 October 2022)
- [13] Knowles 2018 Datasheet of SPH2430HR5H-B (available at: [www.knowles.com/docs/default-source/model-downloads/spw2430hr5h-b-revb.pdf?Status=Master&sfvrsn=bd77b1\\_4](http://www.knowles.com/docs/default-source/model-downloads/spw2430hr5h-b-revb.pdf?Status=Master&sfvrsn=bd77b1_4)) (Accessed 8 October 2022)
- [14] Scheeper P R, Nordstrand B, Gullv J O, Liu B, Clausen T, Midjord L and Storgaard-Larsen T 2003 A new measurement microphone based on MEMS technology *J. Microelectromech. Syst.* **12** 880–91
- [15] Lee S S, Ried R P and White R M 1996 Piezoelectric cantilever microphone and microspeaker *J. Microelectromech. Syst.* **5** 238–42
- [16] White R D and Grosh K 2005 Microengineered hydromechanical cochlear model *Proc. Natl Acad. Sci.* **102** 1296–301
- [17] Shintaku H, Nakagawa T, Kitagawa D, Tanujaya H, Kawano S and Ito J 2010 Development of piezoelectric acoustic sensor with frequency selectivity for artificial cochlea *Sens. Actuators A* **158** 183–92
- [18] Baumgartel L, Vafanejad A, Chen S J and Kim E S 2013 Resonance-enhanced piezoelectric microphone array for broadband or prefiltered acoustic sensing *J. Microelectromech. Syst.* **22** 107–14
- [19] Shintaku H, Kobayashi T, Zusho K, Kotera H and Kawano S 2013 Wide-range frequency selectivity in an acoustic sensor fabricated using a microbeam array with non-uniform thickness *J. Micromech. Microeng.* **23** 115014
- [20] Liu H, Liu S, Shkel A A, Tang Y and Kim E S 2019 MEMS resonant microphone array for lung sound classification *2019 IEEE Int. Electron Devices Meeting (IEDM)* (IEEE)
- [21] Liu H, Liu S, Shkel A A, Tang Y and Kim E S 2020 Multi-band MEMS resonant microphone array for continuous lung-sound monitoring and classification *2020 IEEE 33rd Int. Conf. on Micro Electro Mechanical Systems (MEMS)* (IEEE)
- [22] Zhao C, Knisely K E, Colesa D J, Pfungst B E, Raphael Y and Grosh K 2019 Voltage readout from a piezoelectric intracochlear acoustic transducer implanted in a living guinea pig *Sci. Rep.* **9** 3711
- [23] Jang J *et al* 2015 A microelectromechanical system artificial basilar membrane based on a piezoelectric cantilever array and its characterization using an animal model *Sci. Rep.* **5** 12447
- [24] Von Békésy G 1989 *Experiments in Hearing* (New York: Acoustical Society of America) Wever Ernest Glen
- [25] Stern R M, Morgan N, Virtanen T, Raj B and Singh R 2012 Features based on auditory physiology and perception *Techniques for Noise Robustness in Automatic Speech Recognition* ed T Virtanen, R Singh and B Raj (New York: Wiley Press) ch 8, pp 193–227
- [26] Liu H 2022 MEMS piezoelectric resonant microphone arrays and their applications *PhD Dissertation* University of Southern California
- [27] Rocha B M *et al* 2019 An open access database for the evaluation of respiratory sound classification algorithms *Physiol. Meas.* **40** 035001

- [28] Lea C, Flynn M D, Vidal R, Reiter A and Hager G D 2017 Temporal convolutional networks for action segmentation and detection *Proc. CVPR* pp 156–65
- [29] Infineon 2020 CYBLE-416045-02 (available at: [www.infineon.com/cms/en/product/wireless-connectivity/airoc-bluetooth-le-bluetooth-multiprotocol/airoc-bluetooth-modules/cyble-416045-02/](http://www.infineon.com/cms/en/product/wireless-connectivity/airoc-bluetooth-le-bluetooth-multiprotocol/airoc-bluetooth-modules/cyble-416045-02/)) (Accessed 5 October 2022)
- [30] Shkel A A and Kim E S 2019 Continuous health monitoring with resonant-microphone-array-based wearable stethoscope *IEEE Sens. J.* **19** 4629–38
- [31] Littrell R and Gagnon R 2016 Piezoelectric MEMS microphones noise sources 2016 *Solid-State, Actuators, and Microsystems Workshop Technical Digest (Hilton Head, SC, USA, May)* pp 258–61